Exploring Weather Trends
Prepared by
Donald Ghazi

Objectives: (1) Analyze local and global temperature data (2) Compare tempererature trends
Tools: (1) Anaconda; Python via Jupyter (2) SQL (3) Google Sheets

Step 1: Using SQL queries to retrieve and edit/store data

There are 3 tables in the database; city_list, city_data, and global_data. I opened the city_list table to see how United States may be named in this dataset (could be America, U.S., USA, etc.) and also to identiy the closest big city near me, which is Long Beach.

```sql
SELECT * FROM city_list
WHERE country = 'United States'
```

After selecting my local city, I went ahead and extracted the data only pertaining to Long Beach from the city_data table and saved it as a CSV file.

```sql
SELECT * FROM city_data
WHERE country = 'United States'
    AND city = 'Long Beach'
```

And did the same for the global_data table (export as a CSV file)

```sql
SELECT * FROM global_data
```

NOTE: I made an error on my first attempt as my line chart showed obvious scaling issues. And the reason is that the Long Beach data starts from year 1849 to year 2013 while the global data starts from year 1750 to year 2015 . This means that my moving averages were also calculated incorrectly as well. In order to fix these issues, I started over and cleaned my data more thoroughly. This allowed me to combine the global_data and the city_data tables together (selecting appropriately) and stored as one CSV file.

Now, by simply at looking at the output of the city_data table...

1) I know that I can eliminate reduancies by getting rid of the country column.
2) I can also fix the moving averages and data visualization errors by selecting the years they both each have data for. Essentially being Long Beach years (1849 - 2013).

Now before I perform any SQL queries for above, I need to combine the Long Beach data with the global data. Since these two tables have a column named "avg_temp", I used the ALTER TABLE statement in order to rename a column in an exisiting table and avoid further errors. This was possible by changing the Database Schema of the SQL Workplace server.

```sql
ALTER TABLE city_data RENAME COLUMN avg_temp to lb_avg_temp;
ALTER TABLE global_data RENAME COLUMN avg_temp to gl_avg_temp
```

This allows me to select the data I care about and producing it as one clean table

```sql
SELECT city_data.year,city_data.city,city_data.lb_avg_temp,
global_data.gl_avg_temp
    FROM city_data,global_data
    WHERE(city_data.year = global_data.year)
    AND (country ='United States' AND city = 'Long Beach')
```

The code above allows me to
1) Combine data w/o extracting the country column from city_data and year column from global_data
2) Make sure that x_values are equal (important for calculating moving averages and data visualization).
3) Create well polished data to analyze (easier to work with).

Step 2: Using Google Sheets to calculate moving averages

1) Upload city_data.csv to Google Sheets
2) Calculate Moving Averages for both Long Beach and Global using AVERAGE()function (to smooth out data)
3) I used 10 Moving Average (*Excel commands shown on top). Moving Temperature Averages for Long Beach is 16.89 (ºC). and 9.56 (ºC). for Global.

A	B	C	D
year	city	lb_avg_temp	gl_avg_temp
1849	Long Beach	16.03	7.98
1850	Long Beach	15.55	7.9
1851	Long Beach	15.66	8.18
1852	Long Beach	16.06	8.1

=AVERAGE(C157:C166)

=AVERAGE(D157:D166)

Step 3: Use Python to Visualize Data (Line Chart)

In [118]:
```python
#importing libraries
import pandas as pd # to load data into Jupyter notebook
import matplotlib.pyplot as plt #to make line chart and visualize data
```

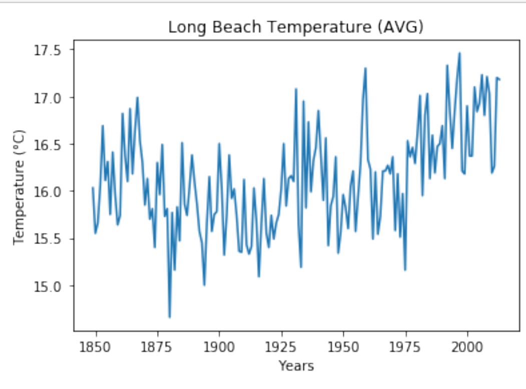
In [119]:
```python
temp = pd.read_csv('clean_data.csv') #importing data set
```

In [120]:
```python
print (temp)
```

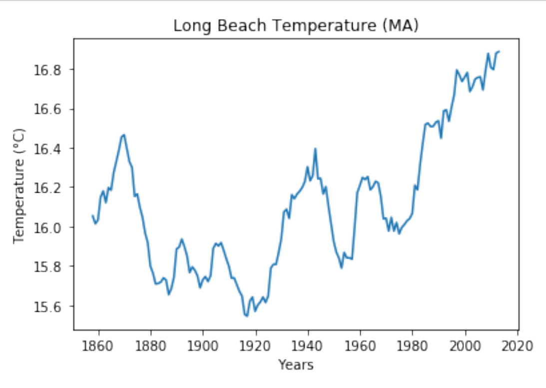
```
     year        city  lb_avg_temp  gl_avg_temp
0    1849  Long Beach        16.03         7.98
1    1850  Long Beach        15.55         7.90
2    1851  Long Beach        15.66         8.18
3    1852  Long Beach        16.06         8.10
4    1853  Long Beach        16.69         8.04
..    ...         ...          ...          ...
160  2009  Long Beach        17.03         9.51
161  2010  Long Beach        16.19         9.70
162  2011  Long Beach        16.26         9.52
163  2012  Long Beach        17.20         9.51
164  2013  Long Beach        17.18         9.61

[165 rows x 4 columns]
```

The table looks good. I proceded to visualize the dataset by first plotting Average Temperature followed by Moving Average Temperature for Long Beach and Global,respectively.

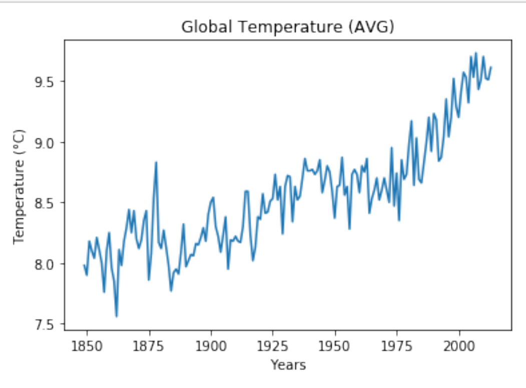
In [121]:
```python
#visualize Long Beach Average Temperature
plt.plot(temp['year'],temp['lb_avg_temp'])
plt.xlabel("Years")
plt.ylabel("Temperature (°C)")
plt.title("Long Beach Temperature (AVG)") #AVG = Average
plt.show()
```



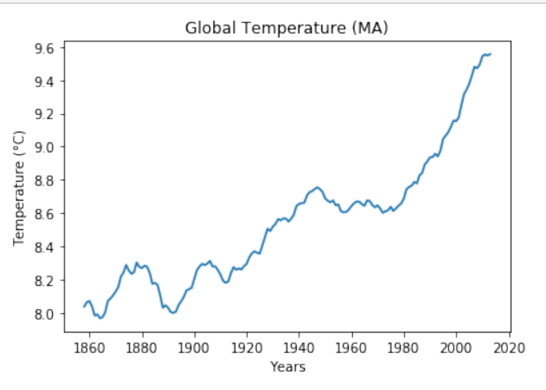
In [122]:
```python
#visualize Long Beach Moving Temperature Average
lb_mv_avg = temp['lb_avg_temp'].rolling(10).mean() # smooth out data using pandas in-built rolling function
plt.plot(temp['year'],lb_mv_avg)
plt.xlabel("Years")
plt.ylabel("Temperature (°C)")
plt.title("Long Beach Temperature (MA)") #MA= Moving Average
plt.show()
```



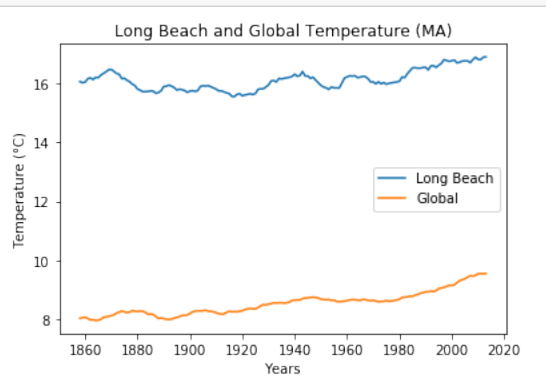
In [123]:
```python
#visualize Global Average Temperature
plt.plot(temp['year'],temp['gl_avg_temp'])
plt.xlabel("Years")
plt.ylabel("Temperature (°C)")
plt.title("Global Temperature (AVG)")
plt.show()
```



In [124]:
```python
#visualize Global Moving Temperature Average
gl_mv_avg = temp['gl_avg_temp'].rolling(10).mean() #10 year moving average (refer to Step 2 above)
plt.plot(temp['year'],gl_mv_avg)
plt.xlabel("Years")
plt.ylabel("Temperature (°C)")
plt.title("Global Temperature (MA)")
plt.show()
```



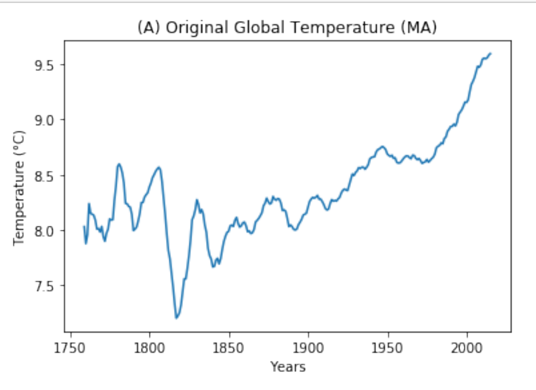
In [125]:
```python
#plot final line chart
plt.plot(temp['year'],lb_mv_avg,label='Long Beach')
plt.plot(temp['year'],gl_mv_avg,label='Global')
plt.xlabel("Years")
plt.ylabel("Temperature (°C)")
plt.title("Long Beach and Global Temperature (MA)")
plt.legend()
plt.show()
```

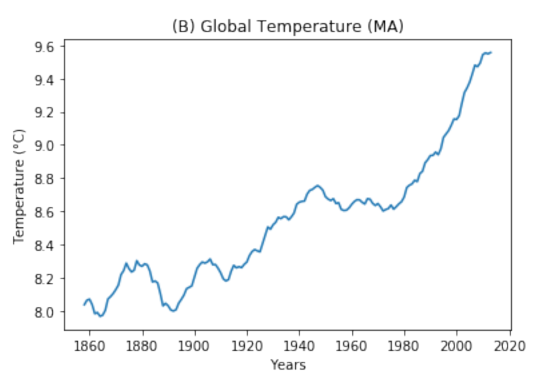


Results

1)Long Beach's weather has been always 2 times hotter than the global average. This is shown on the line chart and can be fact checked when you calculate the Moving Average between year 1849 and 2013 (144 years) which comes out to be 16.12(ºC) for Long Beach and 8.55(ºC) globally.
2) The difference has been consistent over time as they follow similar patterns in temperature change above. Referrring back to 10 year Moving Temperature Average for both, the data show that they are both getting hotter; 16.89 (ºC) and 9.56 (ºC) for Long Beach and the World, respectively.
3)Long Beach's weather has a Mediterranean climate and is near the ocean. However it lacks adequate rain and is home to many industrial plants, 2nd largest cotainer port in the world, and an airport. Naturally, these industries create emissions like nitrogen oxide, which then turn into smog. In 2019, Long Beach and Los Angeles shared the title for having the "most polluted ozone" so I expect the difference between the two to stay consistent for now but the gap between the two may minimize as there are many developing countries becoming hubs to factories.
4) The overall trend for both is that their temperatures are rising. Long Beach is definitely getting hotter and so is the world. Again, comparing the recent 10-year Moving Temperature Averages to 164-year's show that global warming is in full effect. For example, the difference between 164-year and 10-year MA for Long Beach is 4.7% while globally, the difference is 11.8%. This shows that global warming is in full effect and not due to some natural variability.

In [128]:
```python
global_data = pd.read_csv('results-4.csv')
org_gl_ma = global_data['avg_temp'].rolling(10).mean()
plt.plot(global_data['year'],org_gl_ma)
plt.xlabel("Years")
plt.ylabel("Temperature (°C)")
plt.title("(A) Original Global Temperature (MA)")
plt.show()

gl_mv_avg = temp['gl_avg_temp'].rolling(10).mean() #10 year moving average (refer to Step 2 above)
plt.plot(temp['year'],gl_mv_avg)
plt.xlabel("Years")
plt.ylabel("Temperature (°C)")
plt.title("(B) Global Temperature (MA)")
plt.show()
```





5)The charts above uses the original dataset (in order to analyze the world's temperature trend over the few hundred years. Graph A shows the trend of the last few hundred years has been consistent, indicating that the world is getting hotter. It's important to note that drastic spikes in Figure A is due to scaling as there's more x-values in Figure B.